

日 本 国 特 許 庁

PATENT OFFICE
JAPANESE GOVERNMENT

1993 U.S. PRO
09/801285
03/07/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application:

2000年 3月 9日

出 願 番 号
Application Number:

特願2000-071081

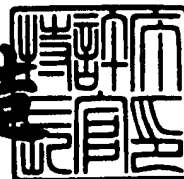
出 願 人
Applicant (s):

ソニー株式会社

2000年12月15日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3104321

【書類名】 特許願

【整理番号】 9900627403

【提出日】 平成12年 3月 9日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 松本 淳

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 西口 正之

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100094053

【弁理士】

【氏名又は名称】 佐藤 隆久

【手数料の表示】

【予納台帳番号】 014890

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9707389

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声信号処理装置及びその信号処理方法

【特許請求の範囲】

【請求項 1】

前向き予測を行う符号化信号の復号処理において、予測残差信号に基づいて音声信号を再生する音声信号処理装置であって、

上記予測残差信号に対して時間軸上で伸長または短縮処理を行う励振源変換手段と、

上記励振源変換手段によって変換された予測残差信号に基づき、音声信号を合成する合成手段と

を有する音声信号処理装置。

【請求項 2】

上記励振源変換手段は、上記予測残差信号を短縮させる処理を行うとき、フレームごとに入力された予測残差信号に対して、前のフレームの予測残差信号から 1 ピッチの長さ L (L は整数、かつ $L > 0$) のうち、 m (m は整数、かつ $m < L$) の信号を取り出し、残り ($L - m$) の信号を参照信号として、次のフレームの予測残差信号から、上記参照信号に最も近似的な信号を切り出し、上記前のフレームから切り出した m の信号の後ろに接続して、1 ピッチ分の新しい予測残差信号を生成する

請求項 1 記載の音声信号処理装置。

【請求項 3】

上記励振源変換手段は、上記次のフレームの信号に対して、上記参照信号との相互相関値を算出し、算出された相互相関値が最も大きい位置から信号を切り出して、上記前のフレームから切り出した m の信号の後ろに接続する

請求項 2 記載の音声信号処理装置。

【請求項 4】

上記励振源変換手段は、上記次のフレームの信号に対して、上記参照信号との二乗誤差を算出し、算出された二乗誤差が最も小さい位置から信号を切り出して、上記前のフレームから切り出した m の信号の後ろに接続する

請求項 2 記載の音声信号処理装置。

【請求項 5】

上記励振源変換手段は、上記予測残差信号を一定の伸長率で伸長させる処理を行うとき、フレームごとに入力された予測残差信号に対して、上記伸長率に応じて、フレームの末尾から逆上って所定の長さのデータを切り出して、上記フレームの後ろに接続して、新しい予測残差信号を生成する

請求項 1 記載の音声信号処理装置。

【請求項 6】

上記合成手段は、線形予測符号化合成フィルタによって構成されている

請求項 1 記載の音声信号処理装置。

【請求項 7】

前向き予測を行う符号化信号の復号処理において、予測残差信号に基づいて音声信号を再生する音声信号処理装置であって、

フレームごとに入力された予測残差信号に対して、ピッチを保持しながら前のフレームの信号と後続のフレームの信号の相互相関に基づき信号を切り出し、また、フレームの信号に対して、外挿法によって推定したデータを接続することによって、1 フレームの信号を短縮または伸長させる励振源変換手段と、

上記励振源変換手段によって変換された予測残差信号に基づき、音声信号を合成する合成手段と

を有する音声信号処理装置。

【請求項 8】

上記励振源変換手段は、上記予測残差信号を短縮させる処理を行うとき、フレームごとに入力された予測残差信号に対して、前のフレームの予測残差信号から 1 ピッチの長さ L (L は整数、かつ $L > 0$) のうち、 m (m は整数、かつ $m < L$) の信号を取り出し、残り ($L - m$) の信号を参照信号として、次のフレームの予測残差信号から、上記参照信号に最も近似的な信号を切り出し、上記前のフレームから切り出した m の信号の後ろに接続して、1 ピッチ分の新しい予測残差信号を生成する

請求項 7 記載の音声信号処理装置。

【請求項 9】

上記励振源変換手段は、上記参照信号と第 1 の窓関数とを掛け算する第 1 の掛け算手段と、

上記次のフレームから切り出した信号と第 2 の窓関数とを掛け算する第 2 の掛け算手段と、

上記第 1 と第 2 の掛け算手段の演算結果を加算する加算手段とを有し、

上記加算手段の加算結果を上記前のフレームから切り出した m の信号の後ろに接続して、1 ピッチ分の新しい予測残差信号を生成する

請求項 8 記載の音声信号処理装置。

【請求項 1 0】

上記励振源変換手段は、上記次のフレームの信号に対して、上記参照信号との相互相関値を算出し、算出された相互相関値が最も大きい位置から信号を切り出して、上記前のフレームから切り出した m の信号の後ろに接続する

請求項 8 記載の音声信号処理装置。

【請求項 1 1】

上記励振源変換手段は、上記次のフレームの信号に対して、上記参照信号との二乗誤差を算出し、算出された二乗誤差が最も小さい位置から信号を切り出して、上記前のフレームから切り出した m の信号の後ろに接続する

請求項 8 記載の音声信号処理装置。

【請求項 1 2】

上記励振源変換手段は、上記予測残差信号を一定の伸長率で伸長させる処理を行うとき、フレームごとに入力された予測残差信号に対して、上記伸長率に応じて、フレームの末尾から逆上って所定長さのデータを切り出して、上記フレームの後ろに接続して、新しい予測残差信号を生成する

請求項 7 記載の音声信号処理装置。

【請求項 1 3】

上記合成手段は、線形予測符号化合成フィルタによって構成されている

請求項 7 記載の音声信号処理装置。

【請求項 1 4】

前向き予測を行う符号化信号の復号処理において、予測残差信号に基づいて時間軸上で伸長または短縮処理を行う音声信号処理方法であって、

フレームごとに入力された予測残差信号に対して、ピッチを保持しながら前のフレームの信号と後続のフレームの信号の相互相関に基づき信号を切り出し、また、フレームの信号に対して、外挿法によって推定したデータを接続することによって、1フレームの信号を短縮または伸長させる処理と、

上記短縮または伸長された予測残差信号に基づき、音声信号を合成する処理とを有する音声信号処理方法。

【請求項 1 5】

上記予測残差信号を短縮させる処理において、フレームごとに入力された予測残差信号に対して、前のフレームの予測残差信号から1ピッチの長さ L (L は整数、かつ $L > 0$)のうち、 m (m は整数、かつ $m < L$)の信号を取り出し、残り($L - m$)の信号を参照信号として次のフレームの予測残差信号から、上記参照信号に最も近似的な信号を切り出し、上記前のフレームから切り出した m の信号の後ろに接続して、1ピッチ分の新しい予測残差信号を生成する

請求項 1 4 記載の音声信号処理方法。

【請求項 1 6】

上記予測残差信号を短縮させる処理において、上記参照信号と第1の窓関数とを掛け算する第1の掛け算処理と、

上記次のフレームから切り出した信号と第2の窓関数とを掛け算する第2の掛け算処理と、

上記第1と第2の掛け算処理の演算結果を加算する加算処理と、

上記加算手段の加算結果を上記前のフレームから切り出した m の信号の後ろに接続して、1ピッチ分の新しい予測残差信号を生成する処理と

を有する請求項 1 5 記載の音声信号処理方法。

【請求項 1 7】

上記予測残差信号を一定の伸長率で伸長させる処理において、フレームごとに入力された予測残差信号に対して、上記伸長率に応じて、フレームの末尾から逆

上って所定長さのデータを切り出して、上記フレームの後ろに接続して、新しい予測残差信号を生成する

請求項 1 4 記載の音声信号処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声信号に対してピッチを変えことなく再生スピードを変化させ、かつ少ない演算量で再生スピードの変化を容易に実現できる音声信号処理装置及びその信号処理方法に関するものである。

【0002】

【従来の技術】

従来、オーディオ信号（音声信号及び音響信号を含む、以下、便宜上単に音声信号として表記する）のピッチを変えないで再生速度を変換するためには、音声信号に対して広い範囲にわたって相互相関演算を行う必要がある。または、音声信号に対して、自由なパラメータ補間を可能とするような枠組み、即ち、音声信号のパラメトリックな表現を予め計算する必要がある。

【0003】

通常、前向き予測を行う音声符号化に対する復号器として、CELP (Code Excited Linear Prediction: 符号励振線形予測) デコーダがある。図 7 は、CELP デコーダの一構成例を示すブロック図である。図示のように、この CELP デコーダは、適応コードブック (Adaptive Codebook : 適応符号帳)、利得コードブック (Gain Codebook : 利得符号帳) 20、雑音コードブック (Stochastic Codebook : 雑音符号帳) 30、バッファ 40、50、加算回路 60、及び LPC 合成フィルタ (Linear Prediction Code Synthesitic Filter: 線形予測符号化合成フィルタ) 70 によって構成されている。

【0004】

CELP デコーダにおいて、ピッチ成分 $e_a(n)$ 及び雑音成分 $e_s(n)$ をそれぞれ振幅調整した信号を加算することによって、残差信号 $e(n)$ が得られる。当該残差信号 $e(n)$ に応じて、LPC 合成フィルタによって音声信号 S_0

(n) が合成される。

【0005】

【発明が解決しようとする課題】

ところで、上述した従来のCELPなどの前向き予測符号化に対する復号処理装置では、音声信号の時間軸上の変換処理は、演算量が大きく、処理が困難であったという不利益がある。

【0006】

本発明は、かかる事情に鑑みてなされたものであり、その目的は、音声信号のピッチ情報を利用して、連続性を維持しつつ予測残差信号の長さを変化させることによって、音声信号のピッチを変えることなく再生スピードを変換でき、かつ少ない演算量で音声信号の再生スピードの変換を実現可能な音声信号処理装置及びその信号処理方法を提供することにある。

【0007】

【課題を解決するための手段】

上記目的を達成するため、本発明の音声信号処理装置は、前向き予測を行う符号化信号の復号処理において、予測残差信号に基づいて音声信号を再生する音声信号処理装置であって、上記予測残差信号に対して時間軸上で伸長または短縮処理を行う励振源変換手段と、上記励振源変換手段によって変換された予測残差信号に基づき、音声信号を合成する合成手段とを有する。

【0008】

また、本発明の音声信号処理装置は、前向き予測を行う符号化信号の復号処理において、予測残差信号に基づいて音声信号を再生する音声信号処理装置であって、フレームごとに入力された予測残差信号に対して、ピッチを保持しながら前のフレームの信号と後続のフレームの信号の相互相関に基づき信号を切り出し、また、フレームの信号に対して、外挿法によって推定したデータを接続することによって、1フレームの信号を短縮または伸長させる励振源変換手段と、上記励振源変換手段によって変換された予測残差信号に基づき、音声信号を合成する合成手段とを有する。

【0009】

また、本発明では、好適には、上記励振源変換手段は、上記予測残差信号を短縮させる処理を行うとき、フレームごとに入力された予測残差信号に対して、前のフレームの予測残差信号から1ピッチの長さ L (L は整数、かつ $L > 0$)のうち、 m (m は整数、かつ $m < L$)の信号を取り出し、残り ($L - m$)の信号を参照信号として次のフレームの予測残差信号から、上記参照信号に最も近似的な信号を切り出し、上記前のフレームから切り出した m の信号の後ろに接続して、1ピッチ分の新しい予測残差信号を生成する。

【0010】

また、本発明では、好適には、上記励振源変換手段は、上記参照信号と第1の窓関数とを掛け算する第1の掛け算手段と、上記次のフレームから切り出した信号と第2の窓関数とを掛け算する第2の掛け算手段と、上記第1と第2の掛け算手段の演算結果を加算する加算手段とを有し、上記加算手段の加算結果を上記前のフレームから切り出した m の信号の後ろに接続して、1ピッチ分の新しい予測残差信号を生成する。

【0011】

また、本発明では、好適には、上記励振源変換手段は、上記次のフレームの信号に対して、上記参照信号との相互相関値を算出し、算出された相互相関値が最も大きい位置から信号を切り出して、上記前のフレームから切り出した m の信号の後ろに接続する。

【0012】

また、本発明では、好適には、上記励振源変換手段は、上記次のフレームの信号に対して、上記参照信号との二乗誤差を算出し、算出された二乗誤差が最も小さい位置から信号を切り出して、上記前のフレームから切り出した m の信号の後ろに接続する。

【0013】

また、本発明では、好適には、上記励振源変換手段は、上記予測残差信号を一定の伸長率で伸長させる処理を行うとき、フレームごとに入力された予測残差信号に対して、上記伸長率に応じて、フレームの末尾から逆上って所定長さのデー

タを切り出して、上記フレームの後ろに接続して、新しい予測残差信号を生成する。

【0014】

また、本発明では、好適には、上記合成手段は、線形予測符号化合成フィルタによって構成されている。

【0015】

本発明の音声信号処理方法は、前向き予測を行う符号化信号の復号処理において、例えば、ピッチ成分と雑音成分に応じて生成された予測残差信号に基づいて時間軸上で伸長または短縮処理を行う音声信号処理方法であって、フレームごとに入力された予測残差信号に対して、ピッチを保持しながら前のフレームの信号と後続のフレームの信号の相互相関に基づき信号を切り出し、また、フレームの信号に対して、外挿法によって推定したデータを接続することによって、1フレームの信号を短縮または伸長させる処理と、上記短縮または伸長された予測残差信号に基づき、音声信号を合成する処理とを含む。

【0016】

また、本発明では、好適には、上記予測残差信号を短縮させる処理において、フレームごとに入力された予測残差信号に対して、前のフレームの予測残差信号から1ピッチの長さ L (L は整数、かつ $L > 0$)のうち、 m (m は整数、かつ $m < L$)の信号を取り出し、残り ($L - m$)の信号を参照信号として次のフレームの予測残差信号から、上記参照信号に最も近似的な信号を切り出し、上記前のフレームから切り出した m の信号の後ろに接続して、1ピッチ分の新しい予測残差信号を生成する。

【0017】

また、本発明では、好適には、上記予測残差信号を短縮させる処理において、上記参照信号と第1の窓関数とを掛け算する第1の掛け算処理と、上記次のフレームから切り出した信号と第2の窓関数とを掛け算する第2の掛け算処理と、上記第1と第2の掛け算処理の演算結果を加算する加算処理と、上記加算手段の加算結果を上記前のフレームから切り出した m の信号の後ろに接続して、1ピッチ分の新しい予測残差信号を生成する処理とを有する。

【 0 0 1 8 】

さらに、本発明では、好適には、上記予測残差信号を一定の伸長率で伸長させる処理において、フレームごとに入力された予測残差信号に対して、上記伸長率に応じて、フレームの末尾から逆上って所定長さのデータを切り出して、上記フレームの後ろに接続して、新しい予測残差信号を生成する。

【 0 0 1 9 】

【発明の実施の形態】

通常、音声信号のピッチを変えることなく再生スピードの変換を行うためには、時間軸上の信号処理、例えば、P I C O L A と呼ばれる処理方法、あるいは周波数軸上におけるパラメータの補間方法の変更による方法がある。本発明は、時間軸における信号処理によって、特に音声信号領域ではなく、残差信号領域における信号処理の方法及びそれを実現するための信号処理装置を提案する。

【 0 0 2 0 】

図 1 は本発明に係る信号処理装置の一実施形態を示す回路図である。

図示のように、本実施形態の信号処理装置は、適応コードブック (Adaptive Codebook : 適応符号帳)、利得コードブック (Gain Codebook : 利得符号帳) 20、雑音コードブック (Stochastic Codebook : 雑音符号帳) 30、バッファ 40、50、加算回路 60、L P C 合成フィルタ (Linear Prediction Code Synthetic Filter: 線形予測符号化合成フィルタ) 70、及び励振源変換回路 (Excitation Modifier) 80 によって構成されている。

【 0 0 2 1 】

図示のように、本発明の音声信号処理装置は、C E L P (Code Excited Linear Prediction: 符号励振線形予測) デコーダに適用したものであり、通常の C E L P デコーダに励振源変換回路 80 を追加したものである。

本発明の音声信号処理装置では、C E L P デコーダにおいて、ピッチ成分 $e_a(n)$ と雑音成分 $e_s(n)$ に応じて算出された残差信号 $e(n)$ に応じて、励振源変換回路 80 によって、データの切り出し、あるいは外挿法などを用いて時間軸上で短縮あるいは伸長処理を行うことによって、ピッチ成分を変化させることなく、音声信号の時間軸上の長さを変化させ、音声信号の再生スピード変換を

実現できる。

【0022】

本実施形態の音声信号処理装置において、適応コードブック10は、入力されたピッチ成分のインデクス S_a に応じて、現在のピッチ成分を示す信号（以下、便宜上単にピッチ成分と表記する） $e_a(n)$ を算出し、バッファ40に出力する。なお、図1に示すように、加算回路60によって算出された残差信号 $e(n)$ が適応コードブック10にフィードバックされる。即ち、適応コードブック10における更新は、通常のコデックと同様に、フィードバックされた残差信号 $e(n)$ に応じて行われる。

【0023】

雑音コードブック30は、入力された雑音成分のインデクス S_p に応じて、現在の雑音成分を示す信号（以下、便宜上単に雑音成分と表記する） $e_s(n)$ を算出し、バッファ50に出力する。

利得コードブック20は、入力された利得のインデクス S_g に応じて、ピッチ成分利得制御信号 g_a 及び雑音成分利得制御信号 g_s をそれぞれ算出し、それぞれバッファ40と50に出力する。

【0024】

バッファ40は、ピッチ成分利得制御信号 g_a によって設定された利得でピッチ成分 $e_a(n)$ の振幅を制御し、ピッチ成分 $e_{a1}(n)$ を加算回路60に供給する。

バッファ50は、雑音成分利得制御信号 g_s によって設定された利得で雑音成分 $e_s(n)$ の振幅を制御し、雑音成分 $e_{s1}(n)$ を加算回路60に供給する。

即ち、ピッチ成分 $e_a(n)$ 及び雑音成分 $e_s(n)$ がそれぞれ利得コードブック20から得られたピッチ成分利得制御信号 g_a 及び雑音成分利得制御信号 g_s によって振幅が制御され、得られたピッチ成分 $e_{a1}(n)$ 及び雑音成分 $e_{s1}(n)$ が加算回路60に送られる。

加算回路60によって、ピッチ成分 $e_{a1}(n)$ と雑音成分 $e_{s1}(n)$ とを加算することによって、残差信号 $e(n)$ を算出して励振源変換回路80に出力する。

【 0 0 2 5 】

励振源変換回路 8 0 は、残差信号 $e(n)$ に対して、切り出しまたは外挿法などの補間処理によって、時間軸上の伸長及び短縮処理を行う。これによって、ピッチを変えることなく時間軸上で長さが変換された残差信号 $e_c(n)$ が求まる。励振源変換回路 8 0 によって得られた残差信号 $e_c(n)$ が駆動音源として LPC 合成フィルタ 7 0 に出力され、音声信号 $S_0(n)$ が再生される。

LPC 合成フィルタ 7 0 は、励振源変換回路 8 0 によって出力された残差信号 $e_c(n)$ 及び外部から入力された LPC 係数 S_p に応じて、音声信号を合成して復元する。励振源変換回路 8 0 によって、時間軸上で伸長または短縮された残差信号 $e_c(n)$ が供給されるため、LPC 合成フィルタ 7 0 によって合成された音声信号 $S_0(n)$ は、もとの音声信号に比べてピッチが変化せずに時間軸上伸長または短縮された音声再生信号となる。

【 0 0 2 6 】

本発明では、上述した適応コードブック 1 0、利得コードブック 2 0、雑音コードブック 3 0 及び LPC 合成フィルタ 7 0 は、従来の CELP デコーダを構成するものと同じものを用いることができる。本発明における励振源変換回路 8 0 は、切り出しまたは外挿法などの補間処理によって、残差信号 $e(n)$ に対して時間軸上で伸長及び短縮処理を行うものである。

以下、励振源変換回路 8 0 の動作についてさらに詳細に説明し、本発明における音声信号の再生スピード変換処理の原理及び処理方法を明らかにする。

【 0 0 2 7 】

励振源変換回路 8 0 は、残差信号 $e(n)$ を処理対象として、当該残差信号 $e(n)$ に対して時間軸上の伸長または短縮処理を行う。以下、外挿法による残差信号 $e(n)$ の短縮処理、即ち、音声信号の再生スピードを上げるために行われる処理を信号波形例を用いて説明する。

【 0 0 2 8 】

図 2 は、励振源変換回路 8 0 において残差信号 $e(n)$ の伸長処理の原理を示す波形図である。図 2 (a) は、残差信号 $e(n)$ の一波形例を示す図である。ここで、音声信号処理装置において、残差信号 $e(n)$ は所定のサンプリング周

波数でデジタル化された信号とする。サンプリング周波数 f_s は、例えば、8 kHz である。音声信号の線形予測符号化 (LPC) 処理では、音声信号に対して時間軸上で分割されたフレーム単位で処理が行われる。1 フレームは、例えば、20 ms の長さを持ち、8 kHz でサンプリングを行う場合、1 フレームにおいて 160 のサンプルデータが得られる。さらに、本実施形態の励振源変換回路 80 における処理においては、各フレームをそれぞれ 4 つのサブフレームに分割し、各サブフレームは、40 サンプルデータを有し、時間軸上 5 ms の長さを持つ。

【0029】

以下、上述した条件において、図 2 (a) に示す残差信号 $e(n)$ に対する短縮処理 (切り出し処理) について説明する。ここで、時間軸上で残差信号 $e(n)$ をもとの長さの半分に圧縮する、即ち、再生スピードを 2 倍に速める場合の処理を例に説明する。

CELP デコーダにおいて、音声信号に対してフォワード予測 (前向き予測) を行うことによって音声信号のピッチが求められる。即ち、励振源変換回路 80 において切り出し処理を行う場合、ピッチが既知である。

【0030】

ここで、フレーム F の間の残差信号を $e(n)$ ($n=0, 1, 2, \dots, 159$) とする。音声信号のピッチの長さを L とする。フレーム F においては、ピッチ L が既知で、ここで、 $L=40$ と仮定する。フレーム F は、4 つのサブフレーム f_1, f_2, f_3, f_4 にさらに分割されている。

【0031】

音声信号の再生スピードを 2 倍にすることは、残差信号 $e(n)$ に基づき、ピッチ L が変化せず、時間軸上でもとの残差信号に対して半分の長さを持つ新しい残差信号 $e_c(n)$ を求めることである。これを実現するために、本実施形態の励振源変換回路 80 では、もとの音声信号の残差信号に対して、1 ピッチ分のデータのうち半分のデータを取り出し、残りの半分のデータを参照信号として、この参照信号に最も近似している信号をもとの残差信号における次の 1 ピッチ分のデータから捜して、見つかったデータを前の 1 ピッチ分から取り出したデータと

合わせて、新しい残差データの1ピッチ分とする。このような処理によって、もとの音声信号のピッチを変えずに、かつもとの音声信号の特徴を維持しつつ再生スピードを2倍にした新しい音声信号を再生できる。なお、参照信号との近似の度合いをはかる方法として、相互相関値、二乗誤差値に基づいて判断することができる。即ち、参照信号との相互相関値が最も大きいあるいは参照信号との二乗誤差が最も小さいという判断基準で参照信号と最も近似している信号を見つけることができる。ここで、一例として、参照信号との二乗誤差（あるいは平均二乗誤差）を基準として、この二乗誤差が最も小さい信号は、参照信号に最も近似している信号とする。以下、図2（a）に示す残差信号の波形を例として、本実施形態の音声信号処理の方法についてさらに詳細に説明する。

【0032】

まず、最初のサブフレーム f_1 において、残差信号 $e(0) \sim e(39)$ の適当な位置から、ピッチ L の半分の長さのデータを取り出して、変換後の残差信号 $e_C(0) \sim e_C(19)$ とする。なお、この切り出しの位置は、例えば、最初のサブフレーム f_1 において、残差信号 $e(n)$ のピークが現れる位置の前後に設定することができる。これによって、ピッチの半分の長さを持つ新しい残差信号 $e_C(n)$ の最初の半分が形成された。

【0033】

次に、残差信号 $e_C(n)$ の1ピッチの後半の部分 $e_C(20) \sim e_C(39)$ を求める。なお、音声信号の長さを圧縮し、かつもとの音声信号の特徴を十分保持するため、後半部分の残差信号 $e_C(n)$ は、次のサブフレーム f_2 から求めることが必要である。ここで、最初のサブフレーム f_1 に、残された後半の残差信号 $e(20) \sim e(39)$ を参照信号 $e_{ref}(n)$ として、当該参照信号 $e_{ref}(n)$ に対して、二乗誤差 $E(i)$ が最も小さい部分をサブフレーム f_2 から見つけ出して、この符号列を新しい残差信号 $e_C(n)$ の1ピッチの後半部分、即ち、 $e_C(20) \sim e_C(39)$ にあてはめる。この二乗誤差 $E(i)$ は、次式の演算によって求められる。

【 0 0 3 4 】

【数 1】

$$E(i) = \sum_{n=0}^{L/2-1} (e_{\text{ref}}(n) - x(n+i))^2 \quad \dots (1)$$

【 0 0 3 5 】

式 (1) において、 $e_{\text{ref}}(n) = e(n+20)$ 、 $x(n) = e(n+40)$ ($n=0, 1, 2, \dots, 19$) である。式 (1) に応じて、それぞれの i における誤差 $E(i)$ を求めて、最も $E(i)$ が小さくなるときの値 i_{opt} を求める。即ち、 i_{opt} は、次式によって求められる。

【 0 0 3 6 】

【数 2】

$$\begin{aligned} i_{\text{opt}} &= \operatorname{argmin} E(i) \\ &= \operatorname{argmin} \sum_{n=0}^{L/2-1} (e_{\text{ref}}(n) - x(n+i))^2 \quad \dots (2) \end{aligned}$$

【 0 0 3 7 】

式 (2) において、 argmin は、その後ろの数式が最小値をとるときの i の値を示す演算子である。

算出された i_{opt} に対して、サブフレーム f_2 の先頭より i_{opt} 番目のデータから、20 個のデータを切り出して、新しい残差信号 $e_c(20) \sim e_c(39)$ とする。即ち、サブフレーム f_1 の後半の信号 $e(n)$ を参照信号 $e_{\text{ref}}(n)$ として、当該参照信号 $e_{\text{ref}}(n)$ に最も近似的な信号をサブフレーム f_2 から見つけ出して、生成される新しい残差信号 $e_c(n)$ の 1 ピッチ分の後半になぎ合わせる。

【 0 0 3 8 】

ここで、例えば、式 (2) に基づく計算の結果、 $i_{\text{opt}} = 15$ と仮定する。このため、サブフレーム f_2 から、15 番目の残差信号データから、連続して 20 個のデータを取り出して新しい残差信号 $e_c(n)$ の 1 ピッチの後半部分とする。即ち、 $e_c(20) \sim e_c(39)$ のデータは、それぞれ $e(35) \sim e(54)$ のデータからなる。

以上の処理によって、新しい残差信号の1ピッチ分のデータ、即ち、 $e_C(0) \sim e_C(39)$ が求められる。図2(b)は、このように算出された残差信号 $e_C(n)$ の波形を示している。

【0039】

次に、2ピッチ目の残差信号 $e_C(n)$ ($n=41, 42, \dots, 79$) を求める。まず、残差信号 $e(n)$ の第3のフレーム f_3 の適当の所、例えば、残差信号 $e(n)$ のピークの位置またはその前後の位置から1ピッチの半分に相当する残差信号 $e(n)$ を取り出して、新しい残差信号 $e_C(n)$ における2ピッチ目の最初の半分とする。

【0040】

残差信号 $e(n)$ において、取り出したデータの末尾から、1ピッチ分のデータの半分に相当する残差信号を参照信号 $e_{ref}(n)$ として、この参照信号 $e_{ref}(n)$ と最も近似的なデータをもとの残差信号 $e(n)$ の4つ目のサブフレーム f_4 の中から探す。そして、上述したように、参照信号との近似の度合いをはかる基準として、式(1)に示すように、参照信号と残差信号の二乗誤差を求める。この二乗誤差が最も小さい値となる位置を i_{opt} とすると、 i_{opt} から1ピッチの半分に相当するデータを切り出して、新しい残差信号 $e_C(n)$ の1ピッチの後半部分のデータとする。

【0041】

ここで、1ピッチ分あたりのサンプリングデータ数を L_1 とし、1フレームのデータ数を N とすると、 $i_{opt} + L_1 / 2 > N$ の場合、新しい残差信号 $e_C(n)$ を形成するために、1フレームの残差信号 $e(0) \sim e(N-1)$ のみでは足りず、残差信号 $e(N-1)$ 以降のデータが必要となる。実際の音声信号処理装置において、音声信号がフレーム単位で入力されてくるので、1フレームの音声符号データを処理している間に、次のフレームのデータがまだ用意できていない場合がある。この場合、1フレームを越えた部分のデータを外挿法などの方法によって、現在処理が行われている1フレームのデータから推定することが必要である。

【 0 0 4 2 】

音声データが一定の時間においては、連続性を有することに着目し、フレームの末尾から逆上って、1ピッチ分のデータを推定値として、1フレームの末尾に接続して、不足分を補う方法は、いわゆる外挿法である。図3は、外挿法によって、1フレームの残差信号にデータを補う処理を示す波形図である。

図示のように、外挿法を用いる場合、1フレームのデータの末尾（ $n = N$ の位置）から、1ピッチ分 L_1 を逆上って到達した位置から1ピッチ分 L_1 のデータを切り出して、この L_1 個のデータを1フレームのあとに足して、データの不足分を補う。さらに、必要に応じて、切り出した1ピッチ分のデータをもう一回追加することもできる。

【 0 0 4 3 】

上述した外挿法によって補われたデータ列 $e_x(n)$ （ $n \geq N$ ）は、次式によって表される。

【 0 0 4 4 】

【数3】

$$e_x(n) = e(n + N - L_1) \quad \cdots (3)$$

【 0 0 4 5 】

外挿法によって、1フレームの残差信号 $e(0) \sim e(N)$ に不足が生じた場合、不足分のデータが補われ、これらのデータを用いて、新しい残差信号 $e_c(n)$ が生成することが可能である。

なお、データを外挿する場合、つなぎ目のデータの不連続性を解消するために、つなぎ目のデータの前後の部分に対して窓関数を掛けたデータを重畳加算することが有効である。

【 0 0 4 6 】

上述した残差信号 $e_c(n)$ の再生方法では、1ピッチ分のデータを生成するために、その前半のデータがもとの残差信号の1ピッチ分の前半のデータを用いて生成され、後半のデータがもとの残差信号の1ピッチ分の後半のデータを参照信号として、この参照信号に最も近似している符号列をもとの残差信号の2ピッチ目のデータから探し、最も近似している信号を新しい残差信号の1ピッチ分の

後半のデータとする。そして、参照信号との近似の度合いをはかる基準として、二乗誤差演算が行われ、この二乗誤差が最も小さい信号を探す。即ち、新しい残差信号 $e_C(n)$ における各ピッチのデータは、その前半と後半はそれぞれ異なるピッチ区間のデータをつなぎ合わせたものであり、データのつなぎ目において不連続性を生じることがある。LPC合成フィルタによって残差信号 $e_C(n)$ に基づいて音声信号を再生する場合には、この残差信号の不連続性を幾分低減する効果が得られるが、さらに不連続性を解消するために、後半のデータの開始部分にあたり、参照信号 $e_{ref}(n)$ と切り出した信号にそれぞれ窓関数を掛けたデータを重畳加算したことによって新しい残差信号 $e_C(n)$ を生成する。

【0047】

窓関数としては通常よく使われている三角窓を用いることができる。図4は三角窓を用いて、残差信号データのつなぎ合わせを示す波形図である。

図4(a)は、もとの残差信号 $e(n)$ の波形を示している。同図(b)は、残差信号 $e(n)$ から切り出した1ピッチの半分の符号 $e(0) \sim e(L_1/2 - 1)$ によって形成された新しい残差信号 $e_C(0) \sim e_C(L_1/2 - 1)$ の波形を示している。そして、残差信号 $e(n)$ の1ピッチの後半のデータを参照符号 $e_{ref}(n)$ として、二乗誤差 $E(i)$ を最小とする位置 i_{opt} が算出され、もとの残差信号 $e(n)$ の2ピッチ目の i_{opt} 番目のデータから、 $L_1/2$ 個のデータを切り出す。

【0048】

上述したように、残差信号 $e_C(0) \sim e_C(L_1/2 - 1)$ のあとに、切り出した $L_1/2$ 個のデータを接続することで1ピッチ分の残差信号 $e_C(n)$ を生成可能である。しかし、このような単純なつなぎ合わせによって生成された残差信号 $e_C(n)$ に不連続性を生じることがある。これを改善するために、ここで、参照信号 $e_{ref}(n)$ と切り出した信号に、それぞれ図4(c)に示す三角の窓関数 $T_1(n)$ と $T_2(n)$ を掛けて、掛け算の結果重畳加算して残差信号 $e_C(n)$ の1ピッチ分における後半のデータとする。図4(d)は、三角の窓関数を用いた重畳加算によって、1ピッチの前半のデータと後半のデータがつなぎ合わせて生成された1ピッチ分の残差信号 $e_C(n)$ の波形を示している。

【0049】

なお、三角の窓関数を掛ける演算処理は、次式に示すように、残差信号の位置に応じた変数 λ を用いて、簡単な掛け算演算によって実現可能である。

【0050】

【数4】

$$e_c(n) = \begin{cases} (1 - \lambda) e_{ref}(n) + \lambda e(i_{opt} + n) \\ \quad \left(\lambda = n / \frac{L}{2}, \right. \\ \quad \left. 0 < n < L/2 \right) \\ e(i_{opt} + n) (L/2 \leq n < N') \end{cases} \quad \dots (4)$$

【0051】

このように、参照信号と切り出した信号にそれぞれ窓関数を掛けて重畳加算して残差信号 $e_c(n)$ を形成することによって、生成される残差信号 $e_c(n)$ のつながり目におけるデータの連続性を改善できる。

【0052】

以上の説明においては、音声信号の再生スピードを上げるための信号処理方法を説明した。音声信号の再生スピードを下げる場合、上述した処理とは逆に、ピッチを変えずに、残差信号 $e(n)$ を時間軸上で伸長する処理を行う必要がある。即ち、残差信号 $e(n)$ に対して、ピッチの長さを保ちながら、例えば、外挿法などの方法によってデータの数を増やす処理を行う。

【0053】

外挿法によってデータの推定を行う場合、音声信号の連続性に着目して、1フレームのデータの末尾から、ピッチの長を単位として1ピッチ分ずつデータを切り出す。そして、切り出したデータ列を1フレーム最後のデータの後ろにつなぎ合わせる。必要があれば、さらに1回目の切り出し位置より1ピッチ分前のデータを切り出して、1回目に外挿したデータの最後尾につなぎ合わせる。

【0054】

図5は、一例として、例えば、もとの音声信号に対して、時間軸上で1.5倍伸長する処理を行う場合、残差信号 $e(n)$ に対して行われる伸長処理を示す波

形図である。

図示のように、この例では、1フレームに4ピッチ分の残差信号データが入っている。即ち、1フレームの長さを N とし、ピッチの長さを L_1 とすると、($N = 4 L_1$)。この残差信号 $e(n)$ に対して、時間軸上で1.5倍を伸長するために、1フレームの符号データに対して、さらに2ピッチ分のデータを増やさなければならない。

【0055】

図5の波形例は、外挿法によって残差信号 $e(n)$ を増やす方法を示している。ここで、1フレームにおいて、4ピッチ分のデータのうち、最後の1ピッチ分のデータを切り出す。そして、フレームの最後尾に切り出したデータ列を2回つなぐ。こうした外挿処理によって、1フレームの N 個のデータ $e(0) \sim e(N-1)$ に対して、さらに2ピッチ分の残差信号 $e(N) \sim e(N+2L_1-1)$ が追加される。即ち、もとの1フレーム分 N 個データに対して、($N+2L_1$)個のデータを含む新しい残差信号 $e_c(n)$ が生成される。残差信号 $e_c(n)$ は、もとの残差信号 $e(n)$ に対して、ピッチの長さが変化していないので、変更後の残差信号 $e_c(n)$ を用いてLPC合成フィルタによって音声信号を生成すれば、ピッチを変えることなく、時間軸上で1.5倍に伸長された音声信号を再生できる。

【0056】

なお、残差信号 $e(n)$ に対する外挿処理は、上述した方法に限られることなく、例えば、図5に示すもとの残差信号 $e(n)$ を時間軸上1.5倍を伸長する場合、もとの1フレーム残差信号 $e(n)$ から、フレームの最後尾から2ピッチ分のデータを切り出して、切り出したデータをフレームの最後につなぎ合わせる。これによって、もとの残差信号に対して、ピッチを変えずに1.5倍に伸長した残差信号 $e_c(n)$ を得られる。この新しい残差信号 $e_c(n)$ をLPC合成フィルタによって音声信号を再生することによって、ピッチを変えることなく、時間軸上で1.5倍に伸長された音声信号を再生できる。

【0057】

なお、上述した外挿法による残差信号データの伸長は、切り出したデータ系列

を単純にもとの残差信号データの後ろにつなぐだけであり、新しい残差信号 $e_c(n)$ のつなぎ目のデータに不連続性が生じることがある。LPCフィルタによって音声信号を再生する場合には、この不連続性を幾分低減する効果が得られるが、さらに不連続性を解消するために、残差信号のつなぎ目のデータに対して、窓関数を用いて、窓関数と掛け算したデータを重畳加算してデータの連続性を向上させることができる。

【0058】

図6は、窓関数として、長さ m の三角の窓関数を用いた接続処理を示している。図6(a)は、残差信号 $e(n)$ の一波形例を示している。図示のように、切り出しのとき、1ピッチ分の長さ L_1 より m ($m < L_1$) 個長いデータ列を切り出す。そして、切り出したデータの先頭の m 個のデータに対して、図6(b)に示す三角の窓関数 $f_1(n)$ を掛ける。一方、もとの1フレームの残差信号 $e(n)$ のデータのうち、最後の m 個のデータに対して、図6(b)に示す三角の窓関数 $f_2(n)$ を掛ける。窓関数を掛けたデータを重畳加算して得られたデータを、残差信号 $e(n)$ のフレームの最後尾から、 m 個前の位置に接続して、その後ろに切り出したデータ列の最初の m 個のデータに続いた L_1 個のデータを接続する。

上述したように1フレーム分のデータの後ろに1ピッチ分のデータを外挿することができる。さらに、外挿したデータの後ろに1ピッチ分のデータを接続する場合に、上述した方法同様に、窓関数を掛け算したデータを重畳加算して接続すればよい。

【0059】

上述したように、三角の窓関数を用いて切り出したデータの先頭及び1フレームのデータの後ろの所定数のデータに対して、窓関数を掛けて、掛け算の結果を重畳加算して新しい残差信号 $e_c(n)$ のデータとして接続することによって、単純な切り出し及びつなぎ合わせにより生じたデータの不連続性を抑制でき、残差信号 $e_c(n)$ に基づいてLPC合成フィルタによって再生された音声信号の連続性の向上を実現できる。

【0060】

以上説明したように、本実施形態によれば、ピッチ情報を保ちながら残差信号を時間軸上で短縮または伸長することによって、生成した新しい残差信号に基づき、LPC合成フィルタで音声信号を合成することによって、ピッチを変えることなく、時間軸上で圧縮または伸長した音声信号を再生でき、即ち、ピッチを変えることなく音声信号の再生スピードを上げたり下げたりすることができる。

【0061】

なお、上述した実施形態は、本発明をCELPデコーダに適用した一例を示した例である。本発明の音声信号再生スピードの変換処理は、CELPデコーダに適用する用途に限られず、音声信号のピッチ情報を含む残差信号を取り扱う他の音声信号処理装置にも同様な原理に基づき適用できることはいうまでもない。

【0062】

【発明の効果】

以上説明したように、本発明の音声信号処理装置及びその処理方法によれば、音声信号のピッチを変えることなく、音声信号の再生スピードを任意に変換することが可能である。

さらに、外挿法などを用いてデータのつなぎ合わせを行う場合、接続位置の前後のデータに対して窓関数を掛けたデータを重畳加算することによって、接続されたデータのつなぎ目の不連続性を改善でき、再生された音声信号の連続性を保ち、音質の向上が図れる利点がある。

【図面の簡単な説明】

【図1】

本発明に係る音声信号処理の一実施形態を示す回路図である。

【図2】

残差信号 $e(n)$ を時間軸上短縮させる場合の処理を示す波形図である。

【図3】

残差信号 $e(n)$ を外挿法によってデータを拡張する処理を示す波形図である。

【図 4】

窓関数を用いて、接続する残差信号のデータ連続性を改善する処理を示す波形図である。

【図 5】

外挿法によって残差信号 $e(n)$ を時間軸上伸長する処理を示す波形図である。

【図 6】

窓関数を用いて残差信号を伸長する場合のデータの連続性を改善する方法を示す波形図である。

【図 7】

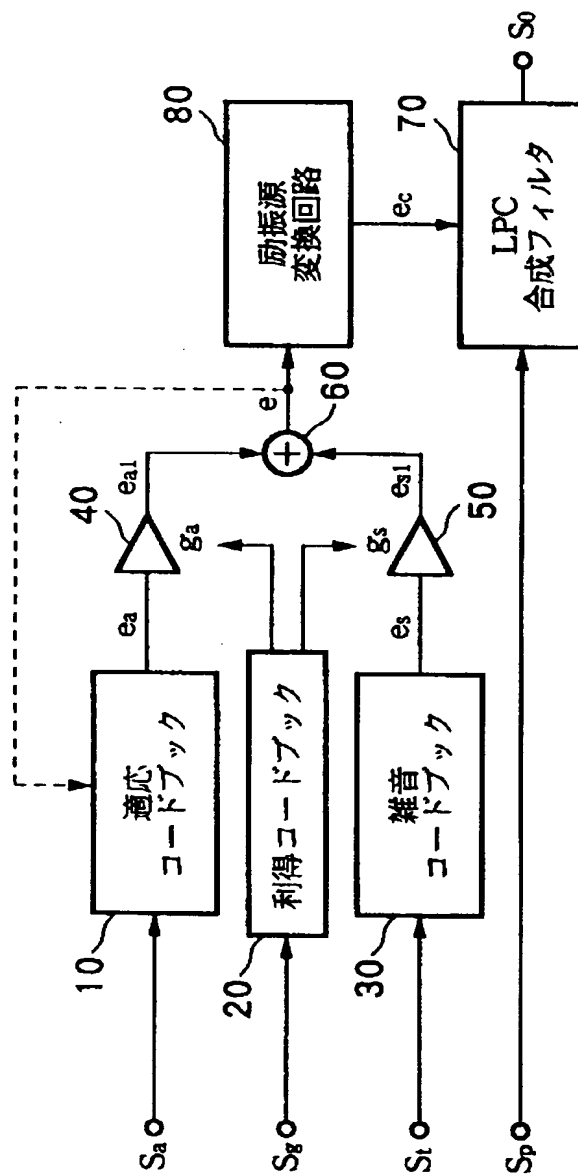
従来の CELP 符号化音声信号復号器の一例を示しブロック図である。

【符号の説明】

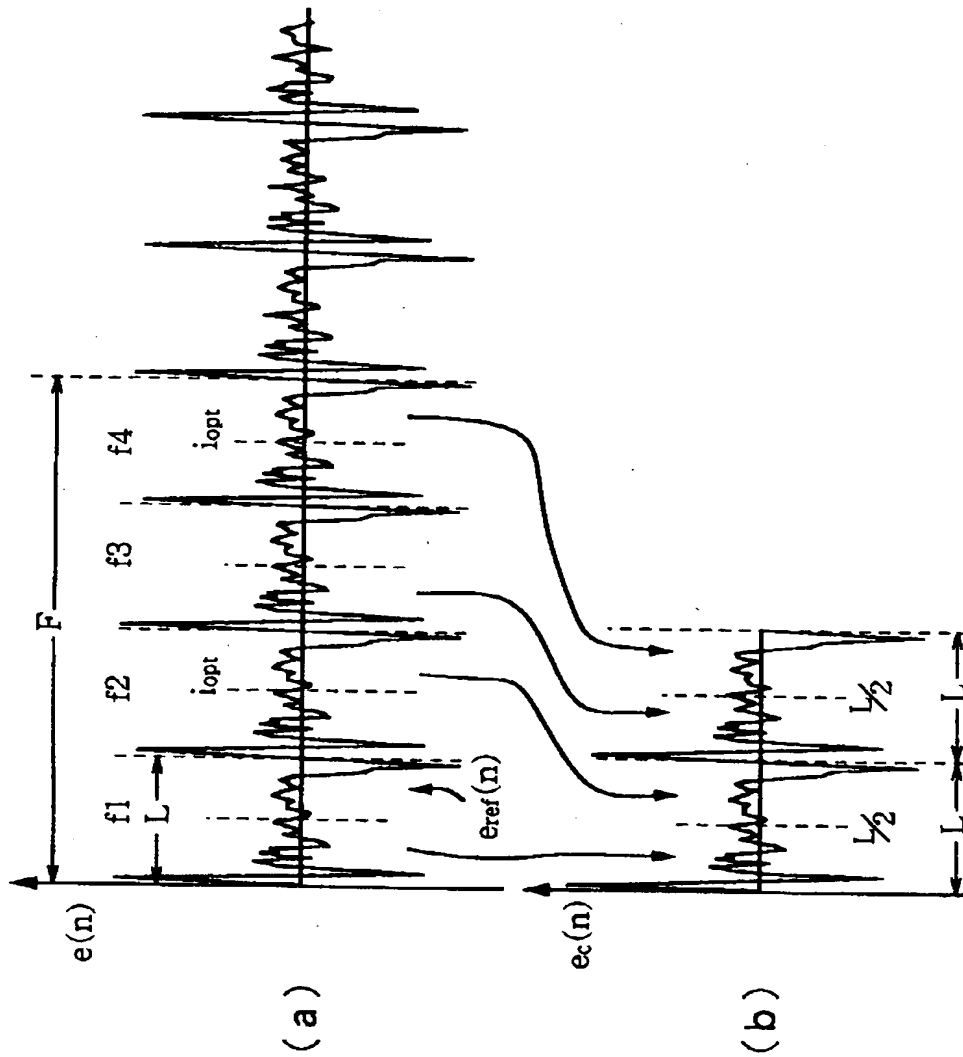
- 10…適応コードブック、
- 20…利得コードブック、
- 30…雑音コードブック、
- 40, 50…バッファ、
- 60…加算回路、
- 70…LPC 合成フィルタ、
- 80…励振源変換回路。

【書類名】 図面

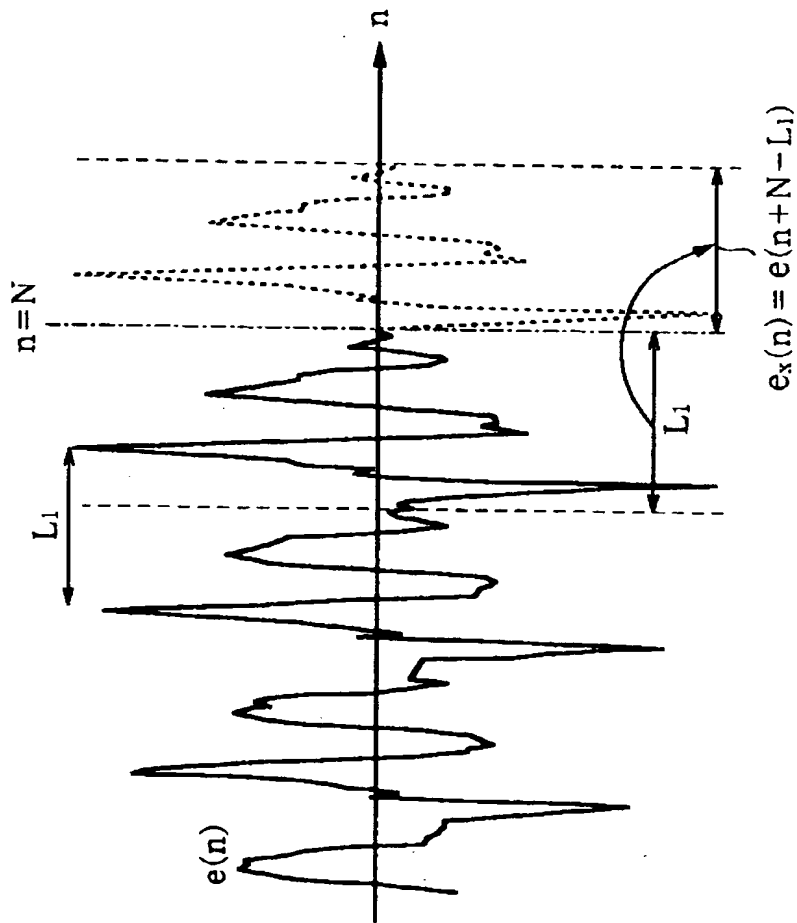
【図 1】



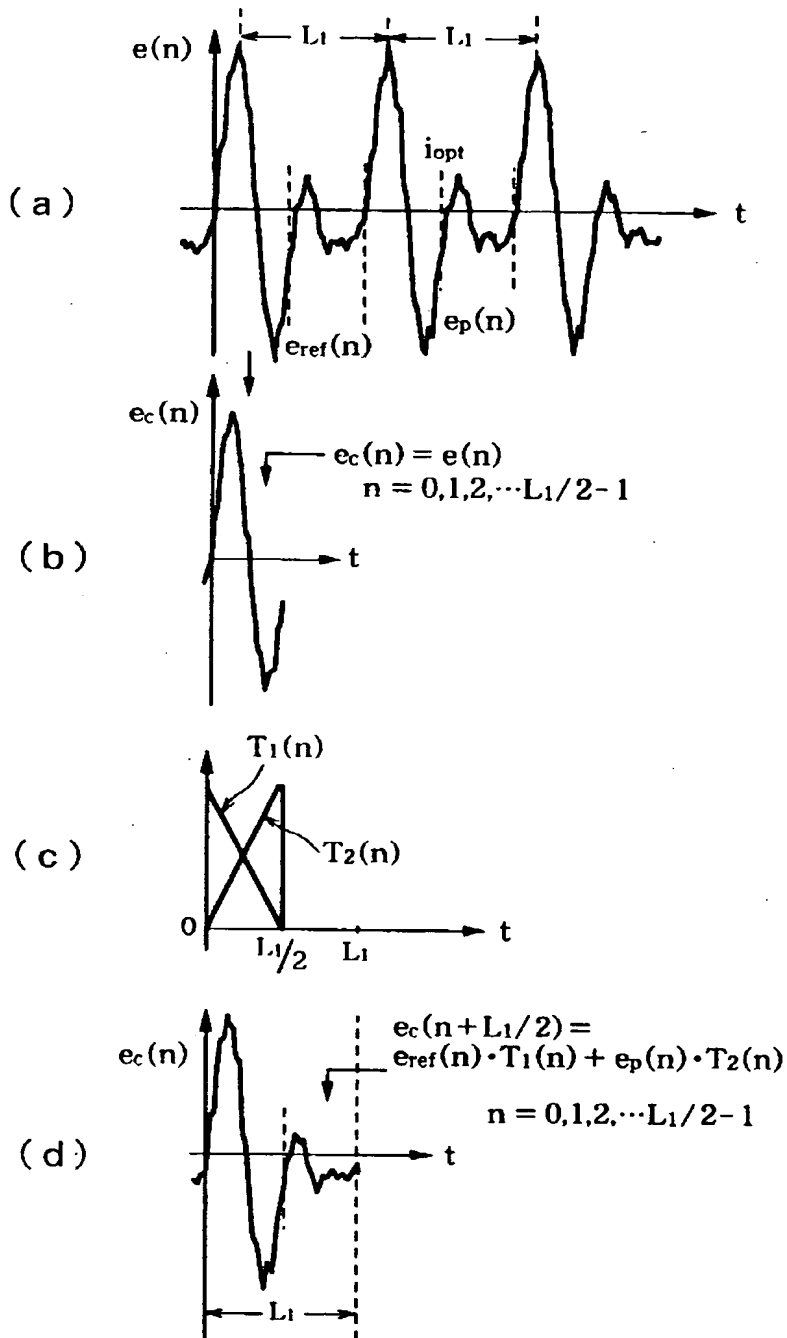
【図 2】



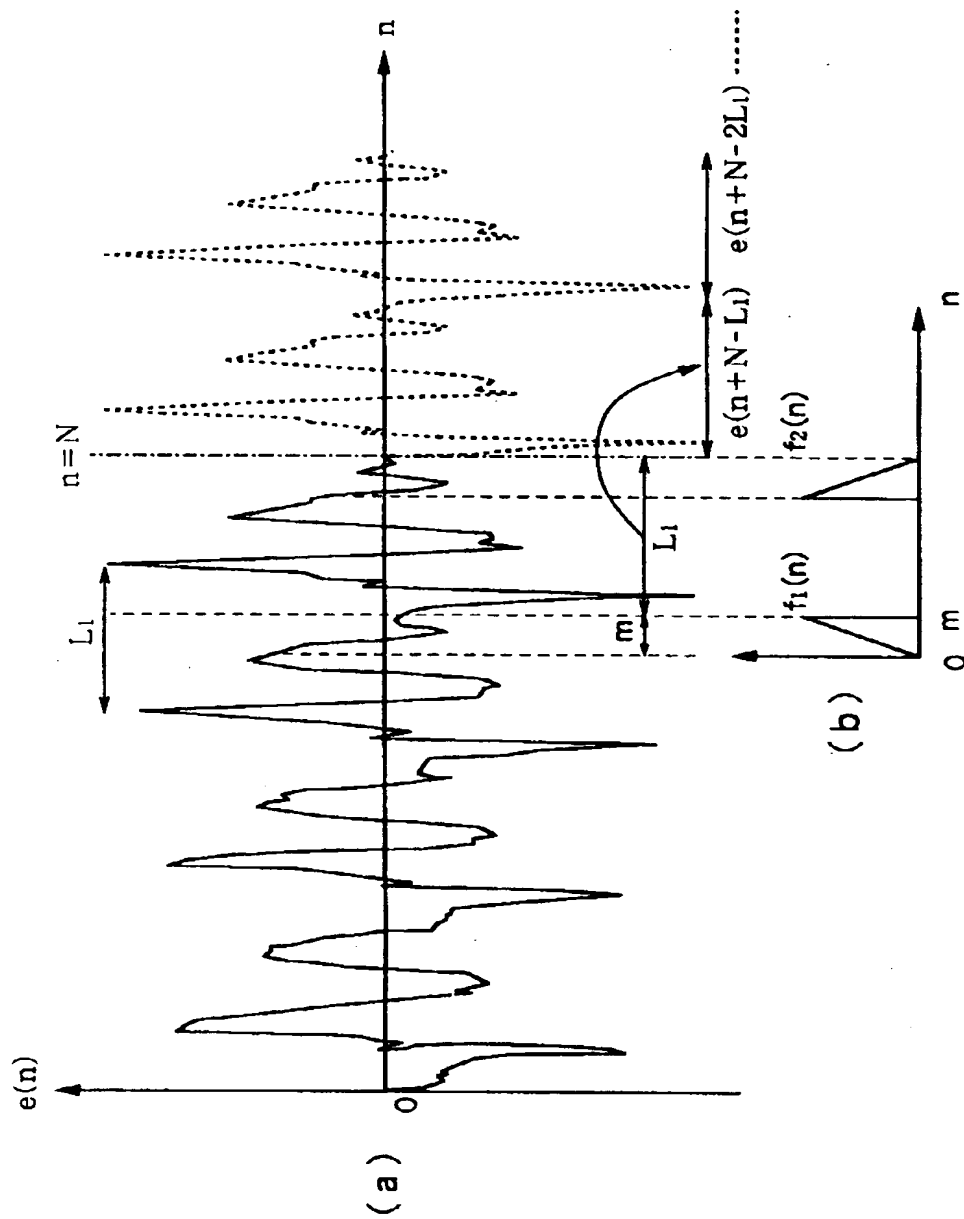
【図 3】



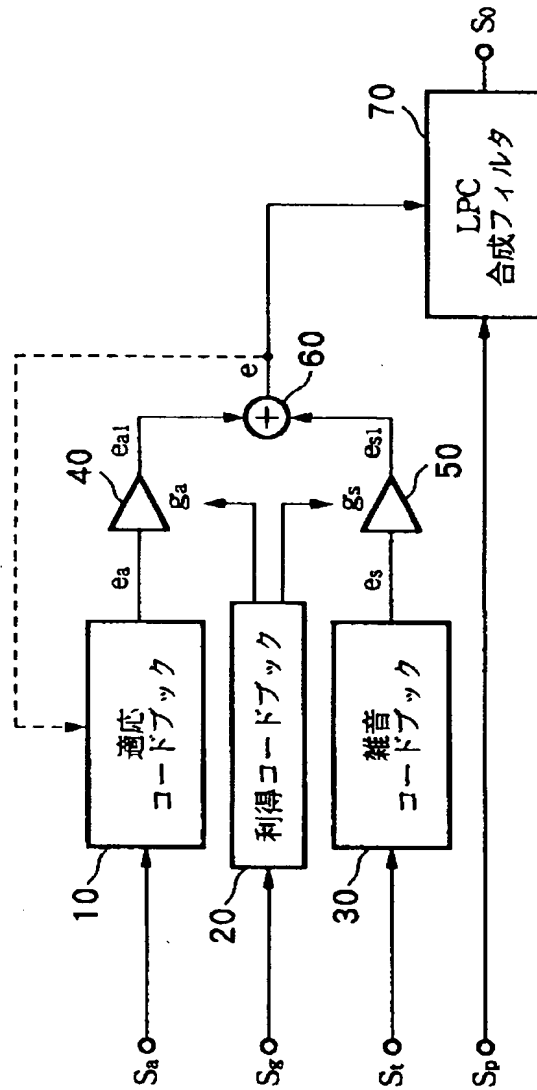
【図 4】



【図 6】



【図 7】



【書類名】 要約書

【要約】

【課題】 ピッチ情報を利用して、連続性を維持しつつ予測残差信号の長さを変化させることによって、ピッチを変えることなく再生スピードを変換でき、少ない演算量でスピードの変換を実現可能な音声信号処理装置を実現する。

【解決手段】 ピッチ情報を保ちながら予測残差信号を時間軸上にて短縮または伸長し、短縮時に信号の相似性に基づき各フレームにおいて異なるピッチ区間の信号を切り出して接続し、伸長時に外挿法によって各フレームの予測残差信号を拡張する。生成した新しい予測残差信号に基づき、LPC合成フィルタで音声信号を合成することによって、ピッチを変えることなく、時間軸上にて圧縮または伸長した音声信号を再生できる。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000002185]

1. 変更年月日	1990年 8月30日
[変更理由]	新規登録
住 所	東京都品川区北品川6丁目7番35号
氏 名	ソニー株式会社